

# Expert Data Scientist & AI Strategist



## **Dr. Mathew Divine**

Dr. sc. hum.

Master of Science (M.Sc.) Biomedical Engineering

Bachelor of Science (B.Sc.) Physics and Mathematics

Frankfurt Region, Germany

email: [mathew.divine@gmail.com](mailto:mathew.divine@gmail.com)

mobile: +49 151 23537195

LinkedIn: <https://www.linkedin.com/in/dr-mathew-divine/>

Publications: [Google Scholar](#)

# Executive Summary

---

"I am a seasoned Data Scientist with over a 15+ years of experience in both traditional machine learning and generative AI in the biomedical and Pharma domains. With a strong foundation in AI strategy, design thinking facilitation, and qualitative user research, I excel in translating complex data analysis into actionable business insights. My expertise includes developing and integrating cutting-edge AI models into business strategies and technical solutions in the cloud, ensuring high-quality outcomes. I have a proven track record of mentoring teams, conducting training, and driving digital transformation initiatives that result in significant cost savings and operational efficiencies. My approach is grounded in collaboration, fostering a culture of innovation and growth. Fluent in both German and English, I am dedicated to leveraging AI to solve complex problems and deliver measurable results."

## Professional Experience

---

01   2024 — present	<p><b>Data Science Expert and AI Strategist — Freelancer</b></p> <p>I specialize in developing and implementing data-driven strategies for organizations looking to enhance their operational efficiency and profitability. My expertise spans from strategic consulting to hands-on implementation, leveraging modern machine learning frameworks and Python to create innovative solutions with a cloud-first mentality.</p> <p><b>Skills:</b> Strategy Consulting, Requirements Gathering, Coaching and Upskilling <b>Technologies:</b> Python, Machine Learning, Cloud Solutions, SQL, Generative AI</p>
08   2021 — 12   2023	<p><b>Head of Digital Lab (CD&amp;O) — Boehringer-Ingelheim</b></p> <ul style="list-style-type: none"><li>• Coached a team of 8 Digital-Experts and -Designers, developing a digital transformation execution strategy, reducing the time to market by several weeks, increasing revenue</li><li>• Aligned execution strategy across departments with workshops, ensuring business unit relevancy and cross-departmental alignment and acceptance</li><li>• Developed and executed a strategy using ChatGPT for document generation, translation, and redaction, based on business outcomes and ROI, Leading to the onboarding of a new vendor, replacing manual processes and saving significant time.</li><li>• Product Owner for analytics landing page integrating with Veeva Vaults, utilizing LLMs for enhanced search and personalisation, increasing access to dashboards and relevancy of dashboards to business operations</li></ul> <p><b>Skills:</b> Team Leadership, Strategic Planning, Agile Methodologies. <b>Technologies:</b> Generative AI, Llama Index, Azure, AWS, Atlassian</p>
01   2021 — 05   2021	<p><b>Interim Head of Data Science — BI X GmbH</b></p> <ul style="list-style-type: none"><li>• Responsible for coaching and development of self-organized team of 9 data scientists, ensuring quality of data science output across multiple initiatives</li><li>• Coaching Product Owners with BI X leadership Team Members during weekly calls to proactively identified and removed blockers to product development</li><li>• Responsible for staffing new initiatives and internal hiring, while maintaining strict adherence to budget and short time-lines</li></ul> <p><b>Skills:</b> Strategic Coordination, Team Management, Leadership <b>Technologies:</b> JIRA, AWS, Atlassian, Git, Python, Machine Learning</p>

01   2018 – 07   2021	<b>(Senior) Data Scientist — BI X GmbH</b> <ul style="list-style-type: none"> <li>Developed a clinical data catalogue integrating historic clinical trials, identifying new biomarkers using NLP and Ontologies.</li> <li>Created ETL processes and search algorithms for a drug discovery app, enhancing collaboration and data quality for new drug compounds.</li> <li>Created PoC and initiated innovation project based upon generation of novel molecules using SMILES and genAI for interactive compound lead optimization web app</li> <li>Lead developer on NGS validation pipeline in AWS for comparing short-read sequencing to long-read assembly methods, resulting in 16 genetic leads for specific disease profiles</li> </ul> <b>Skills:</b> Collaborative Problem Solving, SCRUM Leadership, UX Research <b>Technologies:</b> Python, Docker, SQL, SpaCy, TensorFlow, git, CI/CD
04   2016 – 12   2017	<b>Research Scientist, Radiogenomics, Eberhard Karls University Tübingen</b> <ul style="list-style-type: none"> <li>Aggregated and analyzed high-dimensional datasets including ngs, medical imaging, and gene annotations in clinical trials, allowing physicians to make better treatment decisions</li> <li>Advised students and taught python for Biologists course at the Max Planck</li> </ul> <b>Skills:</b> Interdisciplinary Collaboration, Data Engineering. <b>Technologies:</b> Python, DICOM, TensorFlow, Seaborn, Docker, SQL, NoSQL
03   2010 – 03   2016	<b>Research Fellow, Oncology (WSIC), Eberhard Karls University Tübingen</b> <ul style="list-style-type: none"> <li>Developed Machine Learning pipelines, integrating PET and MRI data to segment tumor microenvironments and validation strategies with histology</li> <li>Co-Authored a Nature Biomedical Engineering Publication demonstrating the translation potential of the aforementioned prognostic biomarker model</li> </ul> <b>Skills:</b> Analytical Thinking, Problem-Solving, Project Management <b>Technologies:</b> MATLAB, PET, MRI

## Education

---

03   2010 – 03   2016	<b>Doctorate in Molecular Imaging (Dr. sc. hum.)</b> <i>Werner Siemens Imaging Center (WSIC), Eberhard Karls University Tübingen</i>
10   2007 - 02   2010	<b>Masters in Biomedical Engineering (M.Sc.)</b> <i>Rheinisch-Westfälische Technische Hochschule (RWTH), Aachen</i>
09   2002 – 05   2007	<b>Bachelors in Physics and Mathematics (B.Sc.)</b> <i>Midwestern State University (MWSU), Texas, USA</i> <i>Received competitive Scholarship from Midwestern State University</i>

## Further Education

---

06   2022	<b>Casual Inference Workshop</b> – Harvard Medical School
01   2022 – 06   2023	<b>1-on-1 Management Coaching</b> – Consulting Impact, Daniel Zacher
01   2022 – 06   2022	<b>Executive Education</b> – Harvard Medical School
2018   2020   2022	<b>Design Thinking Facilitator Training</b> – WeAreNeon

---

<b>Languages</b>	English — Native Language German — Fluent
------------------	--

<b>Personal Interests</b>	listening to music, playing music with others (bass, electric guitar, and drums), personal fitness (body weight workouts, mountain biking, wind surfing, hiking), cooking, reading
---------------------------	--

## **Data Science Expert & AI Strategist - Freelancer**

**2024-01 - 2024-09**

### *Project Lead of CRM Cleaning and Development of AI first Web-Scraping Technology*

- Led CRM data normalization and cleaning project, with multiple steps visualized as a data score card flow using a Sankey Diagram, aiding in customer understanding and buy-in to iteratively develop pipeline further increasing revenue for client
- Implemented and validated a genAI first web crawling strategy, outlining data quality challenges, ensuring high quality, and scalability on AWS to augment customer CRM data

**Soft Skills:** Project Management, Requirements Gathering, Entrepreneurship

**Hard Skills:** Data Visualization, LLM Prompt Prototyping, Cloud Development

**Technologies:** AWS-Sagemaker, Python, LangChain, Snowflake

## **Head of Digital Lab, CD&O - Boehringer-Ingelheim**

**2021-08 - 2023-12**

### *Digital Execution Strategy Development*

- Coached an international team of eight, internal, digital entrepreneurs within the Digital Transformation CoE.
- Co-Created processes for empowered decision making.
- Upskilled team on design thinking methodologies, including design thinking facilitation and qualitative user research methods.
- Developed digital transformation execution strategy, based on data, people, cloud infrastructure, and budget, along with coaching methodologies, problem statement development, executive buy-in, and sponsorship.

**Soft Skills:** Team Leadership, Strategic Planning, Agile Methodologies, Digital Strategy, Project Management, Qualitative User Research

**Hard Skills:** Jira, Confluence, Mural, Office 360

### *Digital Transformation in Medical Writing*

- Performed process mapping, qualitative analysis of internal processes, resulting in business use case descriptions.
- Used ChatGPT to lead medical writers to do PoC on different uses cases including document generation, redaction, and translation.
- Performed market scouting, leading to short list of vendors and the development of Request for Proposal (RfP) process to decrease time to market by several weeks.

**Soft Skills:** Technical Leadership, Cross-Functional Collaboration, Coaching

**Hard Skills:** LLM Prompt Engineering, GenAI, ChatGPT

### *Lay Document Generation Using RAG System*

- Advised Data Scientist and Medical writer on best practices for development of LLM prompts and prompt chaining along with large pdf document ingestion and vector embeddings for quick prototyping.
- Built a RAG system with Streamlit, which enable medical writers to interactively and semi-automatically create lay summaries of clinical trials, based upon the clinical trial protocol.

**Soft Skills:** Technical Leadership, Cross-Functional Collaboration, Coaching

**Hard Skills:** LLM Prompt Prototyping, Prompt Chaining

**Technologies:** Python, llama index, Azure Cognitive Services, ChatGPT, OpenShift

### *Product Owner of Analytics Landing Page*

- Served as the Product Owner for the analytics landing page, incorporating a recommendation engine and utilizing Large Language Models for enhanced question answering and cognitive search capabilities across operational dashboards.
- Created a product vision, roadmap, groomed the backlog for *ad hoc* visualizations of operational data using LLMs, bespoke data models, and OpenAPI specifications.
- Coordinated with multiple stakeholders from business and other technical teams for expectation management and dependency alignment.

**Soft Skills:** Product Ownership, Innovative Thinking

**Hard Skills:** Data Visualization, LLM Utilization

**Technologies:** LLMs, OpenAPI, Azure Cognitive Services, AWS, Jira, Confluence, git

### **Senior Data Scientist - BI X GmbH**

**2018-01 - 2021-07**

#### *Development of Bioinformatics Pipelines for Genome Assembly Algorithms*

- Led a cross-functional team of developers and scientists in creating validation and innovative bioinformatics pipelines on AWS cloud services.
- The project focused on the development of ultra-fast de novo genome assembly algorithms, leveraging cloud computing for enhanced computational efficiency and accuracy.
- Utilized AWS CDK for infrastructure management and orchestrated various AWS services to streamline the bioinformatics pipeline.

**Soft Skills:** Team Leadership, Project Management, Cross-Functional Collaboration

**Hard Skills:** Bioinformatics, Genome Assembly, Pipeline Development, Cloud Computing

**Technologies:** AWS Batch, AWS IAM, AWS Step Functions, AWS ECS, AWS CodePipeline, Python, Docker, AWS CDK, Jira, Confluence, git

### *Clinical Data Catalogue*

- Developed ETL processes in python on the cloud to produce a clinical data catalogue, integrating clinicaltrials.gov and in-house data for seamless access and aggregation of clinical trial information.
- Utilized Ontologies to normalize clinical data over a twenty-year period.
- Implemented semantic search capabilities, utilizing advanced NLP pipelines and graph data models within a knowledge graph on the cloud.
- My role extended to facilitating the product team's ownership, addressing technical and business challenges, and aligning with internal stakeholders to ensure data governance compliance.

**Soft Skills:** Collaborative Problem Solving, Interdisciplinary Communication, Qualitative User Research, SCRUM Leadership

**Hard Skills:** Data Modeling, Exploratory Data Analysis, Semantic Search Algorithm Development, Biomedical Ontologies, NLP, Graph Database Management, Entity Recognition, BERT Transformer Model, CI/CD

**Technologies:** Python, Jupyter Notebooks, Docker, OpenShift, SPARQL, SQL, Stardog, PostgreSQL, SpaCy, Jira, Confluence, git, Jenkins

### *Interdisciplinary Hackathon Creating a Gamification of NLP Annotation in a Web App*

- Led a 2-day interdisciplinary hackathon project focused on gamifying NLP annotation for entity relationship models.
- Responsibilities included desk research, knowledge resource creation for software developers, UX researchers, and external vendors, and guiding data scientists through NLP pipelines.
- Managed the development of a game that allowed biomedical researchers to validate NLP entity linking models in a gamified approach, enhancing knowledge dissemination in knowledge graphs, gamification, and NLP modeling.

**Soft Skills:** Team Leadership, Interdisciplinary Collaboration, Vendor Relationship Management, Knowledge Dissemination

**Hard Skills:** NLP, Data Visualization, Game Design, Entity Relationship Modeling, Knowledge Graph Utilization

**Technologies:** Python, TensorFlow, Jupyter Notebooks, SpaCy, SPARQL, Stardog, Jira, Confluence, git

### *Conceptualization of Bespoke Knowledge Management System Including PoC 2022-10 – 2023-12*

- Led the conceptualization of a bespoke knowledge management system, conducting extensive qualitative user research with over 60 company personnel to determine value proposition and feature set.
- Collaborated with key stakeholders in design sprints to test features and developed prototypes using open-source NLP algorithms for advanced question answering and topic modeling.

- Recognized by the board of governors for innovation and worked with in-house IT for further project development.

**Soft Skills:** User Research, Stakeholder Engagement, Innovative Thinking, Design Thinking, Collaborative Problem Solving

**Hard Skills:** NLP, Question Answering, Topic Modeling, Prototyping, User-Centered Design

**Technologies:** Python, TensorFlow, Elasticsearch, Jira, Confluence, git

#### **Interim Head of Data Science - BI X GmbH**

**2021-01 - 2021-05**

- Responsible for Staffing of new initiatives and internal hiring.
- Collaboratively coached digital Product Owners with BI X leadership team.
- Coordinated Strategy with Global Head of Data Science.

**Soft Skills:** Team Leadership, Agile Methodologies, Digital Strategy, Project Management

**Hard Skills:** Jira, Confluence, Mural, Office 360

#### **Data Scientist - BI X GmbH**

**2018-01 - 2018-12**

##### *Bespoke Drug Discovery and Market Intelligence Web App*

- Led the Data Science development of a web application for drug discovery and market intelligence, focusing on complex ETL processes using Airflow, advanced NLP for entity recognition, and the development of an advanced search feature utilizing biomedical ontologies.
- Actively participated in user interviews alongside UX researchers, providing expertise on biomedical subject matter to inform user-centric design.

**Soft Skills:** SCRUM Leadership, Team Collaboration, Qualitative User Research, Interdisciplinary Communication, Stakeholder Engagement

**Hard Skills:** ETL, NLP, CI/CD, Biomedical Data Processing, Search Algorithm Development

**Technologies:** Postgres, Neo4j, Python, TensorFlow, Airflow, OpenShift, Docker, Jupyter Notebooks, Jira, Confluence, git, Jenkins

##### *Data Science Hackathon Utilizing Generative AI to Generate Novel Medicinal Molecules*

- Led a 2-day hackathon project focusing on utilizing generative AI to create novel medicinal molecules.
- Responsibilities included conducting desk research, creating knowledge resources for non-biomedical data scientists, and leveraging advanced AI models for molecule generation and property estimation.
- Collaborated closely with the Medicinal Chemistry Department to align efforts and co-create a new digital initiative.

**Soft Skills:** Collaborative Problem Solving, Interdisciplinary Communication, Innovative Thinking, Project Management

**Hard Skills:** Data Visualization, Generative AI Modeling, Molecular Property Estimation, Biomedical Data Analysis

**Technologies:** Python, Docker, TensorFlow, Jupyter Notebooks, t-SNE, Vector Embeddings, Jira, Confluence, git

**Research Scientist, Radiogenomics - Applied Bioinformatics Group, Center for Bioinformatics,  
Eberhard Karls University Tübingen** **2016-04 - 2017-12**

*Curation and Reporting of high-dimensional and heterogeneous NGS and Biomedical Imaging Datasets*

- Collaborating with the Radiology, Nuclear Medicine, Bioinformatics, and the Oncology departments to collect, aggregate and analyze high-dimensional and heterogeneous datasets stemming from an observational clinical trial in hepatocellular carcinoma patients.
- I created workflows for curating NGS data sets with multiple online databases.
- I extracted radiomics features from multi-dimensional imaging data sets and used heatmaps, hierarchical clustering methods, and protein interaction networks to visualize and report insights on the project.
- Created interactive reports for treating physicians and iteratively refined based upon feedback

**Soft Skills:** Interdisciplinary Collaboration, Interdisciplinary Communication, Innovative Problem Solving, Public Speaking

**Hard Skills:** Machine Learning, Spectral Clustering, Data Visualization, Data Engineering, Exploratory Data Analysis

**Technologies:** Python, DICOM, TensorFlow, Seaborn

*Clinical Reporting Strategy of Genetic Variations*

- Curating genetic variations of individual patients with online databases and scientific publications.
- The aggregation algorithms were containerized for interoperability, and were used for creation of genetic variation reports to help oncologists decide about alternative therapy options for patients not responding to the standard of care.

**Soft Skills:** Interdisciplinary Collaboration, Interdisciplinary Communication, Public Speaking, Innovative Problem Solving

**Hard Skills:** Data Engineering, Bioinformatics, Exploratory Data Analysis, Data Modeling

**Technologies:** REST API, Python, Docker, Cloud Computing

*Using Parallel Coordinates to Visualize Hidden Layers in Feed Forward Neural Networks*

- Conceptualized an innovative visualization approach for understanding how feedforward neural networks can be visualized with parallel coordinates, and advised master's student to develop interactive web app.



- This approach belongs to the category of explainable AI and allowed for visualization of hidden layers and separability of classes in hidden layers, leading to an intuitive approach to hyper-parameter estimation.

**Soft Skills:** Coaching, Collaborative Problem Solving, Innovative Problem Solving

**Hard Skills:** Data Visualization, Neural Networks, Machine Learning, API Development

**Technologies:** TensorFlow, Python

*Development of interactive webapp for visualization of images, no-code creation of neural network architecture, training, hyper-parameter optimization*

- Conceptualized an innovative workflow tool to visualize large imaging data sets, interactively create neural convolutional auto-encoder architectures, train neural networks, visualize latent representation, and perform batch processing for hyper-parameter estimation.
- I advised a master's student to develop the corresponding, interactive web app. This is an important tool for no-code development of transfer learning techniques for convolutional neural networks and un-supervised clustering and classification of imaging datasets.

**Soft Skills:** Coaching, Collaborative Problem Solving, Innovative Problem Solving

**Hard Skills:** Data Visualization, Convolutional Neural Networks, Machine Learning, Transfer Learning, Clustering

**Technologies:** Python, TensorFlow

*Utilizing Biological Interaction Networks to Aggregate NGS and Molecular Imaging Data, defining highly disrupted pathways*

- Conceptualized a technique to map molecular imaging and NGS data sets to KEGG protein-protein interaction networks and advised a doctoral student to implement the technique.
- This led to a reduction of the dimensionality of the NGS and imaging features, and allowed for introspection of the indicated molecular pathways, which overlapped for a small cohort of patients.
- This represents a novel technique for combining disparate biomedical data sets and pathways analysis.

**Soft Skills:** Coaching, Collaborative Problem Solving, Innovative Problem Solving

**Hard Skills:** Biological Network Visualization, Network Path Analysis

**Technologies:** Python, KEGG, Cytoscape

**Research Fellow, Oncology, Werner Siemens Imaging Center, Department of Preclinical Imaging and Radiopharmacy, Eberhard Karls University Tübingen      2010-03 - 2016-03**

*Development of Gaussian Mixture Modeling Pipeline for Tumor Microenvironment Analysis*

- Conceived and co-developed a novel Gaussian mixture modeling (GMM) pipeline integrating 18F-FDG PET and diffusion-weighted MRI (DW-MRI) data.
- The project aimed to segment the tumor microenvironment into distinct tissue compartments and monitor their evolution over time.
- Successfully demonstrated the pipeline's ability to segment tumor tissue accurately and establish a linear correlation between ADC and 18F-FDG values, suggesting its potential for advancing disease outcome assessment into precision medicine.

**Soft Skills:** Analytical Thinking, Problem-Solving, Interdisciplinary Research, Scientific Communication, Project Management

**Hard Skills:** Data Analysis, Gaussian Mixture Modeling, PET/MRI Data Integration, Precision Medicine Research, Image Processing, Histopathology, Machine Learning

**Technologies:** MATLAB, Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI)

*Multiparametric MRI Tumor Tissue Segmentation Study*

- Initiated and led a study introducing an unsupervised segmentation method using multiparametric MRI to quantify diverse tumor tissue types.
- The study involved comparing this novel approach with existing segmentation techniques and validating the results against histological samples. Findings indicated that the new method accurately quantified the fractional populations of necrotic, perinecrotic, and viable tumor tissues, aligning closely with histological analyses.
- The implications of this research may significantly improve cancer treatment planning and improve MRI-guided tumor biopsies.

**Soft Skills:** Critical Analysis, Innovative Problem Solving, Cross-Disciplinary Collaboration, Data Interpretation, Research Funding Acquisition

**Hard Skills:** MRI Image Segmentation, Unsupervised Learning Algorithms, Histological Data Validation, Cancer Treatment Planning, Biopsy Procedure Enhancement, Histopathology

**Technologies:** MATLAB, Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI)

*Spectral Clustering for Tumor Heterogeneity Assessment in Dynamic 18F-FDG PET Studies*

- This research project explored spectral clustering (SC) techniques for assessing tumor heterogeneity in dynamic 18F-FDG PET imaging studies.
- The study meticulously validated the SC approach against simulations of tumor tissue types and existing methods like compartmental modeling and SUV segmentation.
- The findings underscored SC's robustness in quantifying tumor heterogeneity and its capability for voxel-level characterization of tumor microenvironments, marking a significant step forward in the field of oncological imaging.

**Soft Skills:** Innovative Research, Analytical Thinking, Scientific Communication, Data Validation, Cross-Disciplinary Collaboration

**Hard Skills:** Spectral Clustering, Tumor Heterogeneity Quantification, Dynamic PET Data Analysis, Simulation and Modeling, Image Segmentation, Machine Learning

**Technologies:** MATLAB, Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI)

*Spatial Characterization of Intratumoural Heterogeneity via PET-MRI and Machine Learning*

- This project unveiled a cutting-edge approach to characterize intratumoural heterogeneity spatially through phenotype-specific, multi-view learning classifiers using dynamic PET and multiparametric MRI data.
- The research demonstrated the classifiers' capability to quantify phenotypic changes from targeted therapies in colon cancer models and align with tumour histology in patients with liver metastases from colorectal cancer.
- The application of this novel method in both preclinical and clinical settings underscores its significance in enhancing precision oncology.

**Soft Skills:** Machine Learning Application, Interdisciplinary Collaboration, Innovative Problem-Solving, Scientific Research Communication

**Hard Skills:** Multimodal Imaging Analysis, Machine Learning Classifier Development, Phenotypic Analysis, Dynamic PET-MRI Integration, Biostatistics

**Technologies:** MATLAB, Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI)